National Defence / Défense nationale

DRES

UNCLASSIFIED

SUFFIELD REPORT

NO. 531

(3)

UNLIMITED DISTRIBUTION

NEURAL NETWORK RECOGNITION AND CLASSIFICATION OF AEROSOL PARTICLE SIZE DISTRIBUTIONS

by

E. Yee and J. Ho

PCN No. 051SP

January 1990

DEFENCE RESEARCH ESTABLISHMENT SUFFIELD : RALSTON : ALBERTA

Canada

UNCLASSIFIED

DEFENCE RESEARCH ESTABLISHMENT SUFFIELD

RALSTON ALBERTA

SUFFIELD REPORT NO. 531

NEURAL NETWORK RECOGNITION AND CLASSIFICATION OF AEROSOL PARTICLE SIZE DISTRIBUTIONS

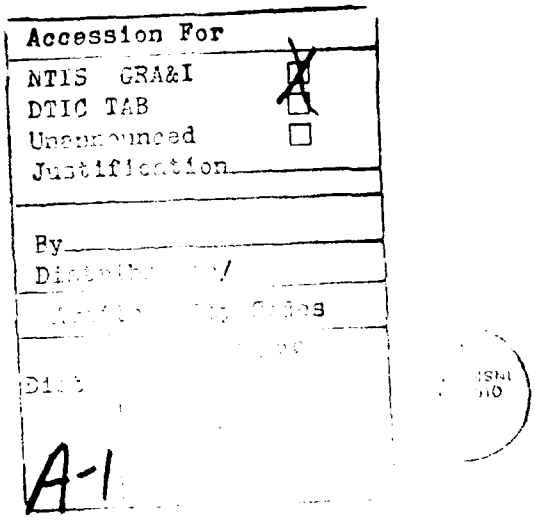
by

E. Yee and J. Ho

PCN No. 051SP

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the valuable comments and encouraging remarks from Dr. T. V. Jacobson, Dr. S. J. Armour and Mr. S. Mellsen on the contents of this paper.



ABSTRACT

This paper describes the application of a neural computational network model to the pattern recognition and classification of aerodynamic particle size distributions associated with a number of environmental, bacterial, and artificial aerosols. The aerodynamic particle size distributions are measured in real-time with high resolution using a two-spot He-Ne laser velocimeter. The technique employed here for the recognition and classification of aerosols of unknown origin is based on a three-layered neural network that has been trained on a training set consisting of 75 particle size distributions obtained from three distinct types of aerosols. The training of the neural network was accomplished with the back-propagation learning algorithm. The effects of the number of processing units in the hidden layer and the level of noise corrupting the training set, the test set, and the connection weights on the learning rate and classification efficiency of the neural network are studied. The ability of the trained network to generalize from the finite number of size distributions in the training set to unknown size distributions obtained from uncertain and unfamiliar environments is investigated. The approach offers the opportunity of recognizing, classifying, and characterizing aerosol particles in real-time according to their aerodynamic particle size spectrum and its high recognition accuracy shows considerable promise for applications to rapid real-time air monitoring in the areas of occupational health and air pollution standards.

EXECUTIVE SUMMARY

In recent years, the military's concern with the enemy use of classical biological agents as well as the newly emerging mid-spectrum agents (which will all be disseminated as solid aerosols) has led to considerable research efforts in the development of aerosol detectors and sensors capable of rapidly identifying and classifying pathogenic aerosols. This paper analyzes and develops a relatively new approach to the adaptive recognition and classification of various types of environmental, bacterial and artificial aerosols that is based on an integrated measurement and computational environment involving the use of an Aerodynamic Particle Size Analyzer for the real-time and high resolution measurement of aerodynamic particle size distributions interfaced with a feed-forward neural network for the adaptive pattern recognition and classification of the observed particle size spectra. The neural network is a computational paradigm based on the concept that a massively parallel network of elemental processors (i.e., artificial neural units) arranged in a manner reminiscent of biological neural nets might be able to learn to recognize and classify patterns in an autonomous manner.

In this paper, it is shown that a fully interconnected three-layered neural network (48 input neurons, a variable number of hidden neurons, and 2 output neurons) with nonlinear sigmoid units for thresholding can be trained with the standard back-propagation learning algorithm using a training set consisting of 25 particle size distribution functions from each of three classes of aerosols (one atmospheric and two latex particle standards). It was found that a recognition rate of 100 percent can be obtained for the training set using neural networks with three or more hidden neurons. Experiments conducted to study the performance characteristics of the neural network as a function of the quality of data used for the training and test sets and of inclusion of random noise in the connection strengths of the trained network showed that the neural network can function as a very fault-tolerant pattern recognition and classification system. Furthermore, it was shown that a fully trained neural network can be used to form reliable generalizations to particle size distributions that it has never "seen". In this regard, it was demonstrated that the trained network was capable of grouping particle size distributions of unknown type into similar categories for the case where the number of cluster categories was unknown *a priori*.

Although the present work has focussed exclusively on the coupling of the neural network paradigm to an Aerodynamic Particle Sizer for aerosol identification, it only represents but one possibility for the recognition and classification of aerosols. It would be desirable to investigate the coupling of adaptive pattern recognition and classification systems based on neural networks (implemented either as software simulations or as electronic and/or optical hardware) with detection methods that provide an optimum sensitivity and specificity for the detection and quantitative evaluation of the chemical or biological agent. In this regard, it might be useful to investigate the application of neural networks to the adaptive recognition and classification of chemical agents (vapors) and of biological agents (biological components in aerosols) based on ion mobility spectrometry and pyrolysis mass spectrometry, respectively.

I. INTRODUCTION

The extraction of features from patterns and waveforms is important in a number of pattern recognition applications spanning a range of different scientific disciplines such as image and scene analysis, computer vision, speech and character recognition, biological taxonomy and medical diagnosis, radar and sonar target recognition and classification, robotics and remote manipulators, and seismic data analysis. The design of an optimal classifier for these patterns and waveforms invariably requires knowledge of the statistics of the signal and noise processes and, with regard to the latter aspect, the performance of the classifer is dependent on how well these statistics have been characterized. However, in most practical applications, the statistics of the signal and noise processes are either not known or have been wrongly assumed to possess certain characteristics. Uncertainties, introduced as the result of unknown noise processes and/or ignorance of the processes and systems generating the waveforms and patterns, require the development of new pattern recognition methods and, in this regard, the self-organizational and perceptual capabilities displayed by artificial neural network architectures offer considerable promise.

The artificial neural network, largely inspired by developments in neurobiology, is a computational paradigm which consists of a network of parallel distributed processing units (i.e., neurons) which are interconnected to one another according to some prescribed topology. Research into artificial neural networks dates back to the seminal work of McCulloch and Pitts [1] in the 1940s and to the development of early two-layered neuronal models in the 1950s and 1960s, such as the PERCEPTRON proposed by Rosenblatt [2] and the ADALINE developed by Widrow [3]. However, interest in neural network research waned in the 1970s after Minsky and Papert [4] demonstrated the limitations and restrictions inherent in all the early two-layered neuronal models. Recently, there has been a resurgence of interest in neural network paradigms and connectionist architectures. This renewed interest has largely been engendered as the result of certain theoretical developments in neural network models and of advances in VLSI technology for the construction and implementation of massively parallel computational architectures [5,6].

The present study applies a neural network computational model to the recognition and classification of environmental (i.e., natural), bacterial, and artificial aerosols on the basis of the aerodynamic particle size distribution. Typical measurements of the aerosol size distributions of atmospheric aerosols with an Aerodynamic Particle Size analyzer indicate that there is a considerable variability in the shape of the particle size spectrum of natural aerosols over a given interval of time and it is this natural temporal variability in shape that makes it difficult to discriminate airborne contamination from natural aerosols using conventional pattern recognition techniques. With regard to the latter point, the detection and classification of sources of contamination from changes in the shape of the background aerosol size distribution is an important process in the adequate assessment of the safety of the environment. Indeed, within the fields of applied and environmental biology, of air quality monitoring, and of toxicological research, the health effects

posed by airborne industrial, bacterial and viral particles depend critically on the ability to recognize, characterize, and classify these particles on the basis of their particle size distribution function. In this paper, we demonstrate how a neural network model may be applied to recognize and classify the complex shapes of particle size distributions of a number of environmental, biological, and artificial aerosols assuming an ignorance of the structural information of the underlying processes generating the aerosols.

The paper is organized as follows. In Section II, a brief description is provided of the architecture of the neural network model used in the present study. Then, in Section III, we review briefly the mathematical formulation of the back-propagation learning algorithm that is utilized to train the network. In Section IV, we describe the aerosol particle size distribution data and the construction of the training and test data sets. Computer simulation results and experiments relating to various aspects of network training, classification performance, fault-tolerance and generalization are presented in Section V. Finally, we discuss our findings and draw some conclusions in Section VI.

II. DESCRIPTION OF NETWORK ARCHITECTURE

The architecture of the neural network used in the present study is illustrated in Fig. 1. The neural network model is essentially a directed graph of processing units or artificial neurons organized into three layers. The input layer of the network consists of 48 neurons—each neuron is associated with one of the 48 aerodynamic particle diameter channels which span the range from 0.5 to 15 μm. Indeed, during normal operation, each of the neurons of the input layer is externally forced or "clamped" to the value of the input aerosol size distribution in a particular particle diameter channel. In this respect, the values of the input neurons are proportional to the probability that the aerosol particle size lies within the particular size interval.

The output layer of the neural network consists of two neurons whose particular state is used to encode the class of the aerosol size distribution. In the present application, the neuron output responses (0,1) and (1,0) are used to represent artificial spherical monodisperse aerosols composed of polystyrene spherical latex particles of 1- and 3-μm diameters, respectively, whereas the output response (1,1) is used to represent natural (atmospheric) aerosol size distributions. Consequently, each of the possible classes of aerosol size distributions is represented by a particular combination of responses of the output neurons. The input neurons are connected to the output neurons through an intervening layer of hidden neurons by a set of connections with adjustable (i.e., adaptive) weights. The connection weights between neurons in the input and hidden layers and the hidden and output layers can be "tuned" by a learning algorithm in order to encode the structural features and higher-order correlations in the input patterns (i.e., aerosol size distributions) that are useful for the detection, identification, and classification of the input. These weights can have both positive and negative values and correspond, respectively, to excitatory and inhibitory connections between the neurons. Finally, it should be noted that the neural network shown in Fig. 1 is feedforward in the sense

that the signal flow in the connections of the network proceeds unidirectionally from the input to the hidden layer and from the hidden to the output layer. In other words, the output of a neuron in a given layer is only connected to the input of the neurons in the subsequent layer.

The structure of a single processing unit or neuron is illustrated in Fig. 2. The output of the i-th neuron is obtained by first computing the weighted sum of the inputs to the neurons according to the prescription

$$S_i = \sum_j w_{ij} p_j , \tag{1}$$

where p_j denotes the output of the j-th neuron and w_{ij} denotes the weight (i.e., interconnection strength) associated with the connection of the output of the j-th neuron to the input of the i-th neuron. The summation in Eq. (1) is over all the inputs to the i-th neuron. The summed output S_i is subsequently subjected to a monotonic sigmoidal transformation which provides a graded response between 0 (minimum) and 1 (maximum). Consequently, the output of the i-th neuron is given by

$$p_i = \phi_\gamma(S_i) = \frac{1}{1 + \exp(-\gamma(S_i + S_i^0))}, \tag{2}$$

where S_i^0 is the threshold or bias for the i-th neuron and γ is the gain of the sigmoid function. The effect of S_i^0 is to shift the sigmoid function to the left ($S_i^0 > 0$) or right ($S_i^0 < 0$) along the horizontal axis, and the effect of γ is to modify the shape of the sigmoid. The sigmoid function with gain $\gamma = 1$ and shift $S_i^0 = 0$ is displayed in Fig. 3, where $X \equiv S_i$ and $Y \equiv p_i$. It is important to note that the soft-limiting sigmoid function serves a threshold device (activation function) for the processing unit and, in this sense, can be considered to be an approximation of the hard-limiting signum function used in some of the early neural network models. However, unlike the signum function, the sigmoid function possesses the necessary differentiability for the application of the back-propagation learning algorithm used for training the neural network.

III. BACK-PROPAGATION LEARNING ALGORITHM

The use of a neural network model consists of two basic phases: (1) a training phase and (2) an operational phase. The primary purpose of the training phase is to encode the features of the input patterns presented to the neural network through the proper selection of the interconnection strengths or weights w_{ij} between the various neurons of the network. To this purpose, a supervised gradient-descent learning scheme known as the back-propagation learning algorithm is utilized to train the neural network model. The back-propagation learning algorithm was developed by Rumelhart *et. al.* [7] and is, in essence, a generalization of the Widrow-Hoff LMS (least mean square) algorithm [8,9] that was originally formulated for adaptive signal processing.

The back-propagation learning algorithm is a gradient-descent algorithm in weight space whereby the output error signals are propagated back through the network in order to modify the weights in the

direction that results in the largest reduction in the error. Application of the back-propagation rule to the training of a neural network involves two passes through the network. In the forward pass, a given training pattern from the training set is presented to the input layer of the network. The signals, generated in the network by the given training pattern, are propagated forwards through the various neurons and their connections according to Eqs. (1) and (2) to produce a response in the neurons of the output layer. In the backward pass, the product of the discrepancy between the observed and desired output responses and the derivative of the threshold (sigmoid) function is propagated in reverse through the connections of the network with the objective of modifying those weights that had a large effect on the output response more than those that did not.

More specifically, the error signal δ_i in the i-th output neuron is calculated according to the prescription

$$\delta_i = (t_i - p_i)\,\phi'_\gamma(S_i), \tag{3}$$

where p_i denotes the response of the i-th output neuron, t_i denotes the desired (i.e., target) response of the i-th output neuron, and ϕ'_γ denotes the first derivative of the threshold function. This output error signal is then back-propagated to the hidden layer where the hidden error signal δ_i^* for the i-th hidden neuron is computed as follows:

$$\delta_i^* = \sum_j \delta_j w_{ij} \phi'_\gamma(S_i), \tag{4}$$

where w_{ij} is the weight associated with the connection from the i-th hidden neuron to the j-th output neuron. Obviously, the summation over j in Eq. (4) is over all neurons of the output layer.

With the calculation of the error signals as per Eqs. (3) and (4), the connection weight w_{ij} between the i-th hidden neuron and the j-th output neuron is modified according to the prescription

$$\Delta w_{ij}(k) = \eta\Big((1-\mu)\delta_j p_i + \mu \Delta w_{ij}(k-1)\Big), \tag{6}$$

where Δ denotes the "change in", η is the learning rate parameter that governs the speed of convergence of the algorithm, μ is the smoothing parameter ($\mu \in [0,1]$), and p_i is the response of the i-th hidden neuron. The index k in Eq. (6) denotes the number of the iteration cycle. Similarly, the weight change in the connection between the i-th input neuron and the j-th hidden neuron is given by

$$\Delta w_{ij}(k) = \eta\Big((1-\mu)\delta_j^* p_i + \mu \Delta w_{ij}(k-1)\Big), \tag{7}$$

where p_i is the response of the i-th input neuron. It should be noted that the smoothing parameter μ serves to suppress oscillations in the weight changes, thus permitting the use of larger values for the learning rate parameter η. In addition to the weights w_{ij}, the thresholds S_i^0 also need to be determined. With regard to this point, it is important to note that these threshold parameters can be determined with the back-propagation learning algorithm in exactly the same manner as for the connection weights. It is only

necessary to imagine the thresholds S_i^0 as the weights from neurons that always have output values of unity. In practice, the elements of the training set are cycled through the neural network and the connection strengths and thresholds are adaptively adjusted with the back-propagation learning algorithm until the discrepancies between all the observed and desired output responses are reduced to below some prescribed tolerance for all the input patterns of the training set.

IV. AEROSOL SIZE DISTRIBUTION DATA

The data sets used for the present investigation are constructed from aerodynamic particle size distribution functions (PSDFs) obtained from 11 different aerosol populations. The PSDFs were measured with an Aerodynamic Particle Sizer (APS), Model 3300 (TSI Incorporated) which determines the aerodynamic diameter of individual aerosol particles by measuring the transit time of the particles between two spots generated by a laser velocimeter that employs a polarized 2-mW He-Ne laser as the light source. The APS brings the aerosol sample into an outer accelerating orifice and focuses the sampled aerosol into an inner nozzle which directs the individual particles through a dual-beam laser formed by splitting a focused laser beam on the basis of polarization using a calcite plate. The beams are then focused using a cylindrical lens to produce two flat beams of rectangular cross-section just downstream of the nozzle orifice. As the aerosol particle passes through these two beams, it triggers a pair of electrical pulses whose temporal separation is accurately measured using a high-speed digital clock. A multi-channel accumulator (MCA) is used to record the transit times of all the aerosol particles and, at the end of a prescribed sampling time, a microcomputer reads each channel of the MCA, translates the channel numbers to aerodynamic particle sizes, and displays the information as a histogram consisting of 48 size intervals (i.e., bins) spanning the 0.5-15 μm aerosol diameter range.

The aerosol size distributions utilized in the study were obtained from artificial, environmental, and biological (i.e., bacterial) aerosols. All the aerosols considered were non-volatile under ambient conditions. The size distributions were classified into 11 categories depending on the source of the aerosol particles generating the distribution. Integer values of 1 through 11 were assigned to these categories for convenient reference. A summary of the aerosol size distribution category notation is found in Table I and a brief description follows. Particle size distributions 1 and 2 correspond to spherical monodisperse polystyrene latex (PSL) particles at nominal 1- and 3-μm diameters and geometric standard deviations (σ_g) of 1.035 and 1.02, respectively. Particle size distribution 3 corresponds to atmospheric aerosols (i.e., background) composed of a population of aerosol particles of both natural and anthropogenic origin suspended in the atmosphere. A mixture of equal proportions of 1- and 3-μm PSL particles, of 1-μm PSL particles and atmospheric aerosols, of 3-μm PSL particles and atmospheric aerosols and, of 1-, 3-μm PSL particles and atmospheric aerosols provide particle size distributions 4, 5, 6, and 7, respectively. Particle size distribution 8 coincides with 0.6-μm PSL particles with a geometric standard deviation of 1.05. Finally, particle size

distributions 9, 10, and 11 correspond to viable rod-shaped aerosolized bacteria *Erwinia herbicola* (EH) (2.5 × 0.5 μm), native *Bacillus subtilis var. globigii* (BG) (1.5 × 0.5 μm), and "clean" *Bacillus subtilus var. globigii* (1.5 × 0.5 μm), respectively. The latter bacterial particles were obtained from the native BG by washing once with ultrapure water followed by centrifugation. The EH and BG provide examples of long-rod and short-rod inhomogeneous bacterial cells, respectively. Furthermore, it should be noted that EH and BG were aerosolized in the form of liquid suspensions of vegetative cells and spores, respectively, and these particles were found to be almost completely dehydrated within a few seconds after aerosolization.

All the aerosol size distributions measured were normalized to take values between 0 and 1 before they were used as input to the neural network. Particle size distributions 1, 2, and 3 were used to train and test the network. To this purpose, the PSDFs in each of these three classes were divided equally to form two sets: a set of 25 PSDFs were randomly selected from each of the three classes to form the training set (i.e., the training set consists of 25 PSDFs from each of the three classes for a total of 75 PSDFs) and the remaining 25 PSDFs from each of the three classes served as the test set. The aerosol size distributions from categories 4 to 11 were used to study the operational properties of the trained neural network and, in particular, to investigate the capability of the trained network to generalize, recognize, and classify PSDFs for which it has not been trained.

V. NEURAL NETWORK EXPERIMENTS

Neural network experiments were carried out using the aerosol data sets described above. The networks were simulated in software on a Compaq 386/20 computer with a neural network simulator program developed by California Scientific Software [10]. The experiments on training of the neural network were performed with the smoothing parameter $\mu = 0.1$ and the learning rate parameter $\eta = 1.0$, unless otherwise indicated. The neural network, which consisted of 48 input neurons in the input layer, 2 output neurons in the output layer, and a variable number of hidden neurons in the hidden layer, was fully interconnected, viz., each neuron in the input layer was connected to every neuron in the hidden layer and, in turn, each neuron in the hidden layer was connected to every neuron in the output layer. Each of the 48 neurons in the input layer was "clamped" to the value of the aerodynamic particle size distribution in one of the corresponding 48 size intervals or bins. Training proceeds by presenting the PSDFs in the training set to the neural network and a training cycle will refer to one presentation of all (75) PSDFs in the training set to the network. The desired output response of the network was chosen so that aerosol size distributions in categories 1, 2, and 3 gave output neuron responses (0,1), (1,0), and (1,1), respectively. These responses identify the three pattern classes in the training set. The connection strengths (weights) in the neural network were initialized with random values drawn from a uniform distribution with a range from −1 to 1. Training of the network with the back-propagation learning algorithm proceeded until the all output responses were within a tolerance of 10 percent of the desired responses.

A set of learning curves which characterize the speed of convergence (i.e., learning speed) of the back-propagation rule for adjusting the weights of the network was desired. One realization of a set of learning curves (percent of the input/output pairs in the training set correctly mapped as a function of the number of training cycles) of the network is shown in Fig. 4 for the specified number of hidden neurons. In this figure, the learning curves are plotted for neural networks with $n = 3$, 5, 7, 10, and 20 hidden units. Observe that the learning rates of the networks generally increase with an increase in the number of hidden units since the use of more hidden units provide a greater flexibility in the network encoding process. Furthermore, although there is a significantly faster learning rate for the network with 10 hidden neurons compared with the network with 3 hidden neurons, this improvement in learning rate does not seem to continue for networks with more than 10 hidden units. Indeed, the networks with 10 and 20 hidden units have almost identical learning behavior. The smooth ensemble average learning curves were obtained by using the average of 20 individual learning curves that correspond to the use of different randomly selected initial weight values. The results are presented in Fig. 5 for networks with $n = 3$, 5, and 10 hidden neurons. The effect of the learning rate parameter η on the learning behavior of the network is exhibited in Fig. 6. Here, the number of training cycles required to fully train the network is plotted against the number of hidden units in the network for two values of the learning rate parameter, namely $\eta = 1.0$ and 1.5. As expected, increasing the learning rate parameter results in an improvement in the learning behavior (i.e., less training cycles are required to train the network) for a specified number of hidden units. However, it is important to emphasize that if η is chosen too large, the back-propagation algorithm exhibits instability and fails to converge properly. Also evident in Fig. 6 is the decrease in the learning rate with an increase in the number of hidden units in the network.

After the networks were fully trained, it was found that input of the PSDFs in the test set to the networks provided recognition accuracies of 100 percent for all the networks trained, viz. for networks with numbers of hidden neurons ranging from 3 to 20. Next, we investigated the effects of noise on the recognition performance of the neural network. All the following experiments involved neural networks with 20 hidden units. It is important that the network be robust in the sense that slight to moderate perturbations in the system (i.e., degradations in the connection strengths of the fully trained network) should not adversely affect the recognition performance. Towards this purpose, consider, for example, the representative result exhibited in Fig. 7. The figure shows the recognition accuracy of the neural network on the test set after all the weights in the trained network had been corrupted with zero-mean Gaussian noise with the standard deviation adjusted to provide the required root-mean-square (RMS) noise level. The RMS noise level is defined by the ratio σ_n^2/σ_s^2 expressed as a percentage, where σ_n^2 and σ_s^2 are the variances in the noise and signal, respectively. Observe that the recognition performance of the network was relatively insensitive to deviations in all the connection weights up to a RMS noise level of about 20 percent. However, after this noise level, the performance of the network deteriorated precipitously with the addition of further noise to

the weights. Furthermore, similar results were obtained when the weights were corrupted with non-Gaussian noise (e.g., a noise process drawn from a uniform distribution). Consequently, the neural network model is relatively fault-tolerant to perturbations in the components of the system.

Fig. 8 exhibits the result of training the neural network with a noise corrupted training set. This figure shows the recognition performance of the network on the test set after training the network on a training set corrupted with noise at the specified level. Observe that the trained network achieved an recognition accuracy of 100 percent for a RMS noise level of up to 10 percent on the training patterns. After this noise level, the performance of the network monotonically decreased (albeit, rather slowly) until it provided a recognition accuracy of 70 percent at the 50 percent RMS noise level. It should be noted that at these higher levels of noise degradation, the input PSDFs are no longer consistent, with the result that the network cannot be trained to classify the training set with an accuracy of 100 percent. In fact, it is observed in these experiments that the recognition accuracy of the trained network on the test set was roughly equal to the recognition performance on the training set. Along the same theme, Fig. 9 depicts the recognition performance of a trained network on the test set that has been degraded with different levels of noise. Note that this performance curve is remarkably similar to that in Fig. 8. Again, the recognition rate of the network remained at 100 percent for up to 10 percent RMS noise level in the test set. Above this level, the performance gradually decreased until a recognition accuracy of only 40 percent was achieved at 55 percent RMS noise level. Along these lines, Figs. 10 and 11 display representative output neuron responses (i.e., activation levels of output neuron 2 versus that of output neuron 1) of the trained network to test sets that have been corrupted with RMS noise at the 5 and 25 percent levels, respectively. Observe that the three categories of PSDFs in the test set produced the expected clustering patterns in the output response space with the scatter within these clusters increasing with the noise level. In summary, the results in Figs. 7 to 11 indicate that the recognition performance of the neural network model is reasonably tolerant to degradations in the connection weights and to variations of PSDFs in the training and test set. We remark that the insensitivity of the neural network to these forms of degradations is not surprising since it is the inherent parallelism and built-in redundancy embodied in the various interconnections of the network that render this form of computational paradigm so attractive for processing of data involving incomplete and/or degraded information.

The next series of experiments investigates the ability of the fully trained neural network to form meaningful generalizations on aerosol size distributions for which it has never been trained. Specifically, we are interested in the ability of the trained network to function as a feature map classifier for a number of unlabeled particle size distributions where the number of categories is unknown *a priori*. In the following experiments, a neural network with 20 hidden units was trained on the training set and then presented with a number of PSDFs that it has never "seen". In Figs. 12 and 13, we show the output network response to a set consisting of a random mixture of 25 PSDFs from each of categories 1 to 6 for a total of 150 PSDFs.

Observe that the PSDFs from classes 1, 2, and 3 (i.e., the classes which the network was trained to recognize) cluster in the expected positions in the output response space. Note that there is no scatter in the responses of the network to PSDFs from classes 1 and 2, indicating that the PSDFs from these classes are highly reproducible. The PSDFs from classes 4, 5 and 6 (i.e., three classes for which the network has not been trained) yielded clusters of points in the output response space with centroids at (0.55,0.35), (0.35,0.95), and (1.0,0.5), respectively. It is interesting to note that the trained network seems to have formed meaningful generalizations from the training examples. In particular, PSDFs from class 4, which are composed of a superposition of aerosols from classes 1 and 2, are clustered roughly about a point in the output response space that lies on the line connecting (0,1) (cluster point for PSDFs of class 1) and (1,0) (cluster point for PSDFs of class 2). A similar statement can be made with respect to the centroids of the clusters for PSDFs of classes 5 and 6 (mixture of aerosols from classes 1 and 3 and 2 and 3, respectively) which roughly lie on the lines connecting the points (0,1) and (1,1) (cluster point for PSDFs of class 3) and (1,0) and (1,1), respectively. Fig. 14 shows the output response of the neural network to a set comprised of PSDFs from class 7 which is a mixture of aerosols from classes 1, 2, and 3. The centroid of the cluster in the response space is located at the point (0.84,0.73). Observe that this point lies within the triangle whose vertices are (0,1), (1,0), and (1,1) which are the centroids of the clusters for PSDFs from classes 1, 2, and 3, respectively.

Next, we study the ability of the trained network to classify bacterial aerosols. Fig. 15 illustrates the network output response to a data set composed of PSDFs from categories 1, 2, 3, 8, and 9, with each category contributing 25 PSDFs. The PSDFs of classes 8 (0.6 μm PSL particles) and 9 (EH) produced clusters in the response space with centroids at (0.38,0.87) and (0.11,0.90), respectively. Only one sample PSDF from class 9 was incorrectly placed. Next, a data set was constructed from PSDFs from classes 3, 9, 10, and 11 and used as input to the trained network. The output responses of the network for this data set are shown in Fig. 16. Observe that the network was able to separate the various classes of PSDFs and place them in distinct clusters in the output response space. Indeed, only 2 PSDFs from the data set were incorrectly placed to provide a classification accuracy of 98 percent. Of interest is the fact that the network was able to recognize the difference in the shape of PSDFs between native and "clean" BG. It is hypothesized that washing the native BG removed extracellular material from the cell wall of the bacteria which resulted in a subtle alteration in the aerodynamic properties of the cell. This alteration produced a subtle change in the shape characteristics of the aerodynamic particle size distribution which was recognized by the trained neural network. With regard to the latter point, the centroids of the clusters formed from PSDFs of classes 10 (native BG) and 11 ("clean" BG) are located at (0.20,0.95) and (0.60,0.87), respectively.

As a final example, an experiment was conducted to study the response of the trained network to an aerosol PSDF that slowly evolves over a given interval of time. Towards this objective, an aerosol of "clean" BG was slowly sprayed into an aerosol chamber and the particle size distribution of the evolving mixture in the chamber was measured every 3 seconds with an APS. A total of 100 PSDFs was measured

over a period of 300 seconds and this suite of PSDFs provided a "spectrogram" of the evolving PSDF in the chamber. A data set was constructed from these 100 PSDFs and 25 PSDFs from category 3. The output responses of the neural network to this data set are exhibited in Fig. 17. Observe that the PSDFs from class 3 are clustered about the expected position. The responses of the network to the PSDFs of the "spectrogram" generally displayed an ordered trajectory of points in the response space that is indicative of an evolutionary behavior. These points eventually clustered about the point (0.60,0.87), which coincides with the centroid of the response cluster for "clean" BG (cf. Fig. 16). This clustering of points in the response space corresponds to that portion of the spectrogram where the PSDFs have reached a steady-state in the aerosol chamber. At this point, the PSDFs no longer change shape as a function of time.

VI. CONCLUSIONS

In this paper, we have applied a neural network model for the recognition and classification of a number of aerosol particles (e.g., environmental, bacterial, and artificial) based on their aerodynamic particle size distribution as measured with an Aerodynamic Particle Size analyzer. It was demonstrated that a fully interconnected three-layered neural network (48 input neurons, a variable number of hidden neurons, and 2 output neurons) with nonlinear sigmoid units for thresholding can be trained with the standard back-propagation learning algorithm using a training set consisting of 25 PSDFs from each of three classes of aerosols (one atmospheric and two polystyrene latex particle standards). It was found that a recognition rate of 100 percent can be obtained for the training set using neural networks with three or more hidden units and that there was a monotonic increase in the learning rate (viz., a smaller number of passes through the training data) with an increase in the number of hidden units in the network. However, it is important to emphasize that there was virtually no increase in the learning times of the networks with more than 10 hidden neurons. Furthermore, it is interesting to note that the performance of the networks did not deteriorate when the number of hidden units was increased beyond 10, despite the fact that for these networks, the number of connection weights that have to be adjusted greatly exceeded the number of training data. In this respect, the back-propagation rule appears to be stable for underdetermined problems. After training, the connection weights were frozen at their final values and a further pass through the test data set consisting of 25 PSDFs from each of the same three classes of aerosols represented in the training set, yielded a 100 percent recognition rate. Evidently, the trained network has properly encoded the significant characteristic features in the PSDFs to permit effective recognition and classification of these three aerosol classes.

Experiments were conducted to study the performance characteristics of the neural network as a function of the quality of data used for the training set and the test set and of the inclusion of random noise in the connection strengths of the trained network. Firstly, it was found that the neural network is structurally robust in the sense that deviations in the connection weights (i.e., up to about 20 percent RMS noise in all

the weights) did not adversely affect the recognition accuracy of a fully trained network. Consequently, the neural network paradigm, with its high degree of parallelism and redundancy in interconnections between neurons, functions as a very fault-tolerant pattern recognition and classification system. Secondly, it was found that the neural network was robust with respect to the type and level of noise corrupting the data in both the training and test sets. In this regard, the neural network was able to perform adequately in the face of uncertainties introduced as the result of undesirable disturbances in the data. This is in contrast to all statistical techniques for pattern classification which are invariably affected by noise type and level. Consequently, the neural network model is more suited for classification of signals from systems where one is confronted with ignorance of the statistical characteristics of the noise corrupting the signals.

After the neural network has been fully trained, it was shown that the network was capable of forming reliable generalizations to PSDFs that it has never "seen". In other words, a properly trained neural network can be used to rapidly characterize and classify airborne environmental, chemical and biological, and artificial aerosols with a high accuracy rate. Indeed, it was demonstrated that a neural network that was trained to recognize and classify only three categories of aerosols can be used effectively to classify aerosols from eight other categories for which it has never been trained. It was shown that the aerosols from these eight categories produced separable clusters in the output response space of the trained network. Consequently, the trained network was shown to be able to group PSDFs of unknown type into similiar categories for the case where the number of cluster categories was unknown *a priori*. Indeed, for a properly trained neural network, cluster classification accuracies of better than 98 percent were obtained for all eight aerosol classes utilized. This rather surprising classification performance on unknown PSDFs indicates that the trained network had properly encoded the characteristic features of the PSDFs of the training set in the connection strengths and had been able to generalize this encoding to extract characteristic features in the unknown PSDFs and to use this information to form appropriate cluster categories for the PSDFs in the output response space. In view of this, a properly trained neural network coupled to an APS provides one possibility of recognizing, characterizing, and classifying aerosol particles in real-time.

REFERENCES

1. McCulloch, W. S. and Pitts, W., "A Logical Calculus of the Ideas Imminent in Nervous Activity", Bull. Math. Biophys., Vol. 5, 115-133, 1943.

2. Rosenblatt, F., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, Spartan Books, Washington, D. C., 1961.

3. Widrow, B., "Generalization and Information Storage in Networks of Adaline Neurons", in Self-Organizing Systems, Yovitz, M.C., Jacobi, G. T. and Goldstein, G. (eds.), Spartan Books, Washington, D. C., 435-461, 1962.

4. Minsky, M and Papert, S., Perceptrons: An Introduction to Computational Geometry, MIT Press, Cambridge, MA, 1969.

5. Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations, MIT Press, Cambridge, MA, 1986.

6. Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models, MIT Press, Cambridge, MA, 1986.

7. Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation", in Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (eds.), MIT Press, Cambridge, MA, 318-362, 1986.

8. Widrow, B. and Hoff, M.E., "Adaptive Switching Circuits", in IRE WESCON Conv. Rec., pt. 4, 96-104, 1960.

9. Widrow, B. and Stearns, S. D., Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1985.

10. BRAINMAKER: Users Guide and Reference Manual (3rd Edition), California Scientific Software, Sierra Madré, CA, 1989.

TABLE I

Summary of class notation used for aerosol size distributions.

Aerosol particle descriptor	Numerical assignment
PSL particles (1 μm)	1
PSL particles (3 μm)	2
Atmospheric	3
PSL particles (1 and 3 μm)	4
PSL particles (1 μm) and atmospheric	5
PSL particles (3 μm) and atmospheric	6
PSL particles (1 and 3 μm) and atmospheric	7
PSL particles (0.6 μm)	8
Erwinia herbicola (2.5 × 0.5 μm)	9
Native *Bacillus subtilus var. globigii* (1.5 × 0.5 μm)	10
"Clean" *Bacillus subtilus var. globigii* (1.5 × 0.5 μm)	11

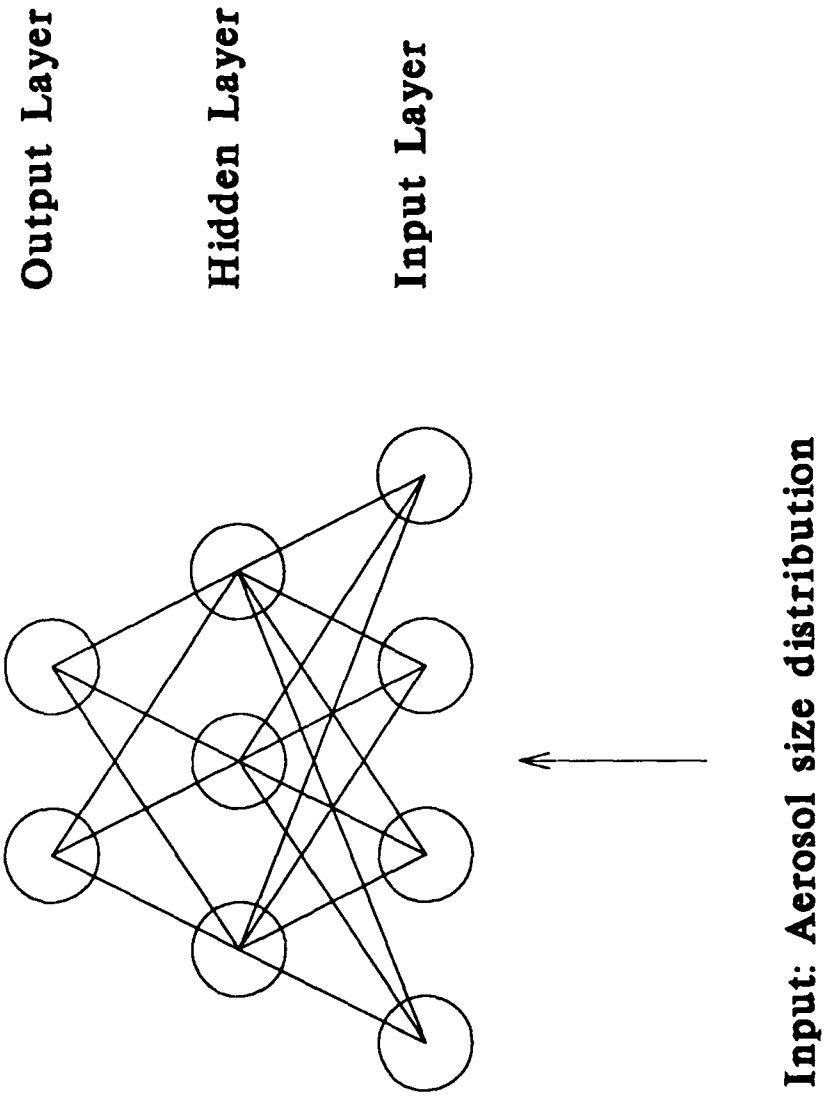


FIGURE 1

Schematic of the architecture of a feedforward three-layered neural network.

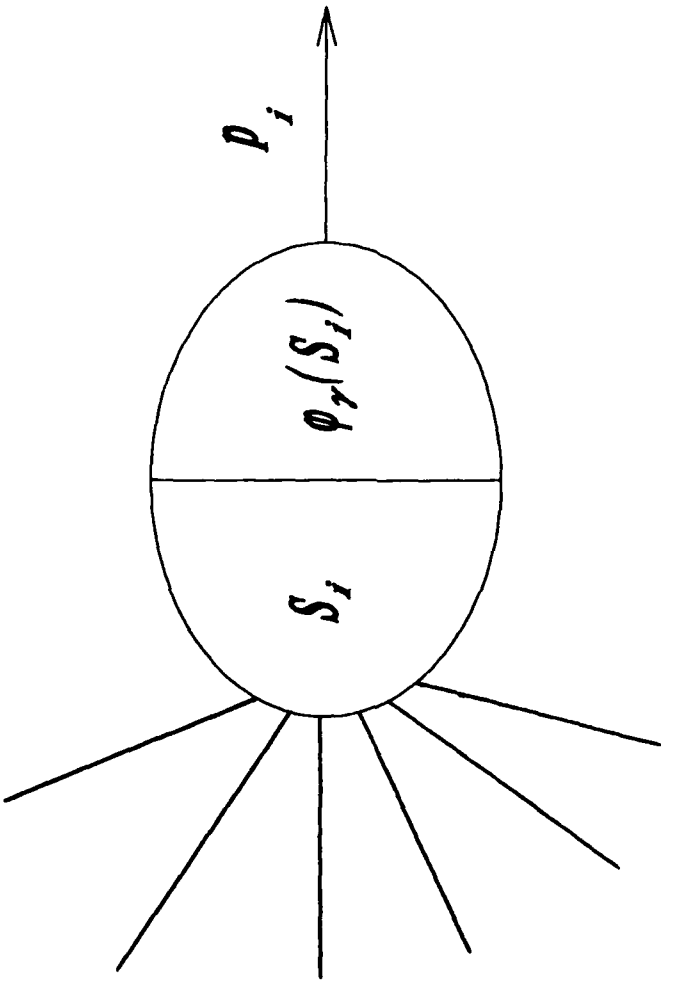


FIGURE 2

The components of a single processing unit (neuron).

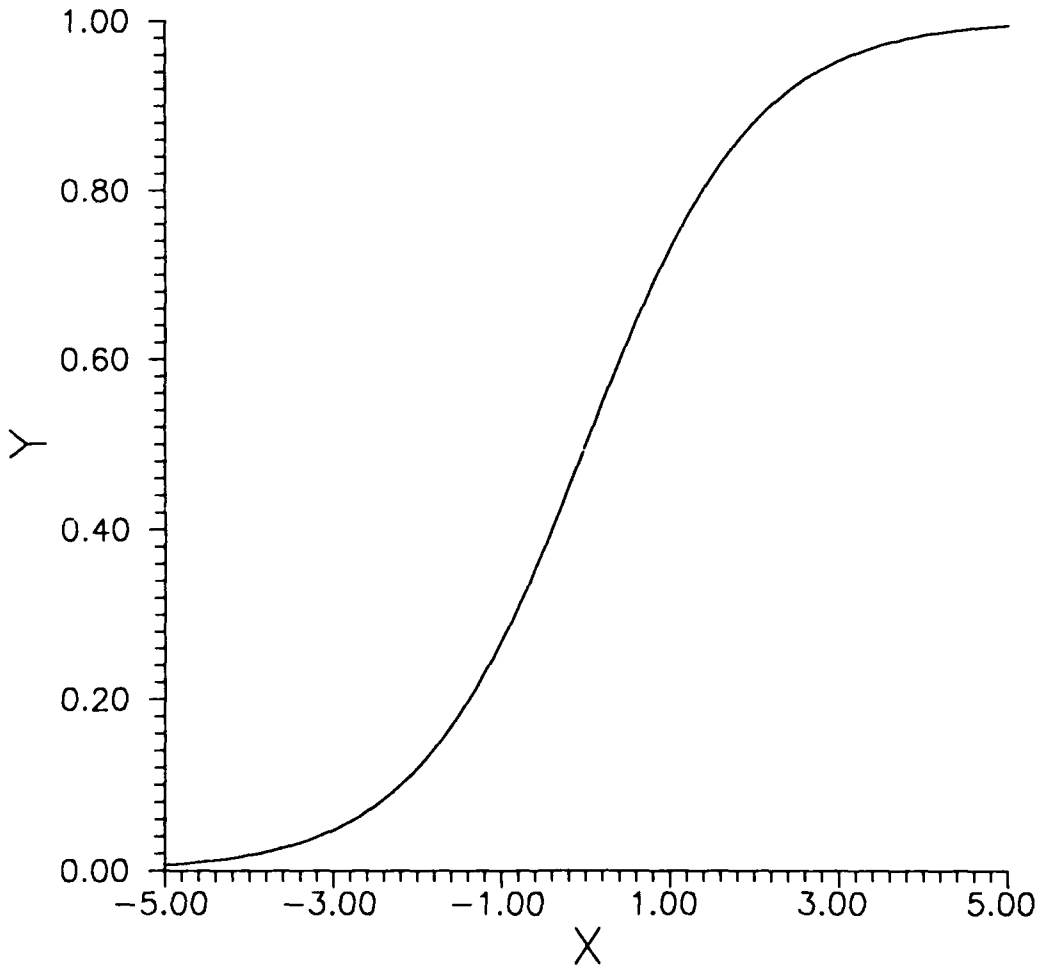


FIGURE 3

The sigmoid function with unit gain that serves as the threshold or activation function for a neuron.

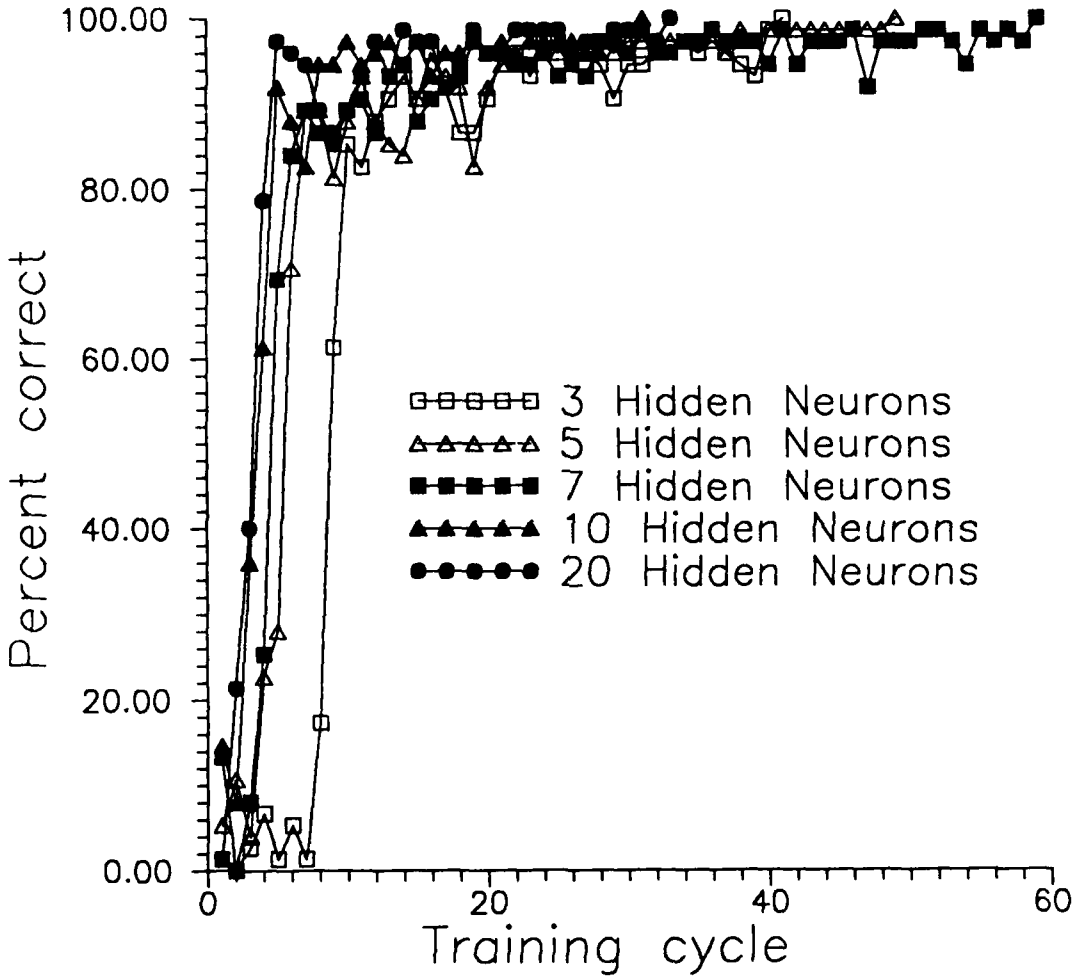


FIGURE 4

One realization of learning curves for neural networks with 3, 5, 7, 10, and 20 hidden neurons.

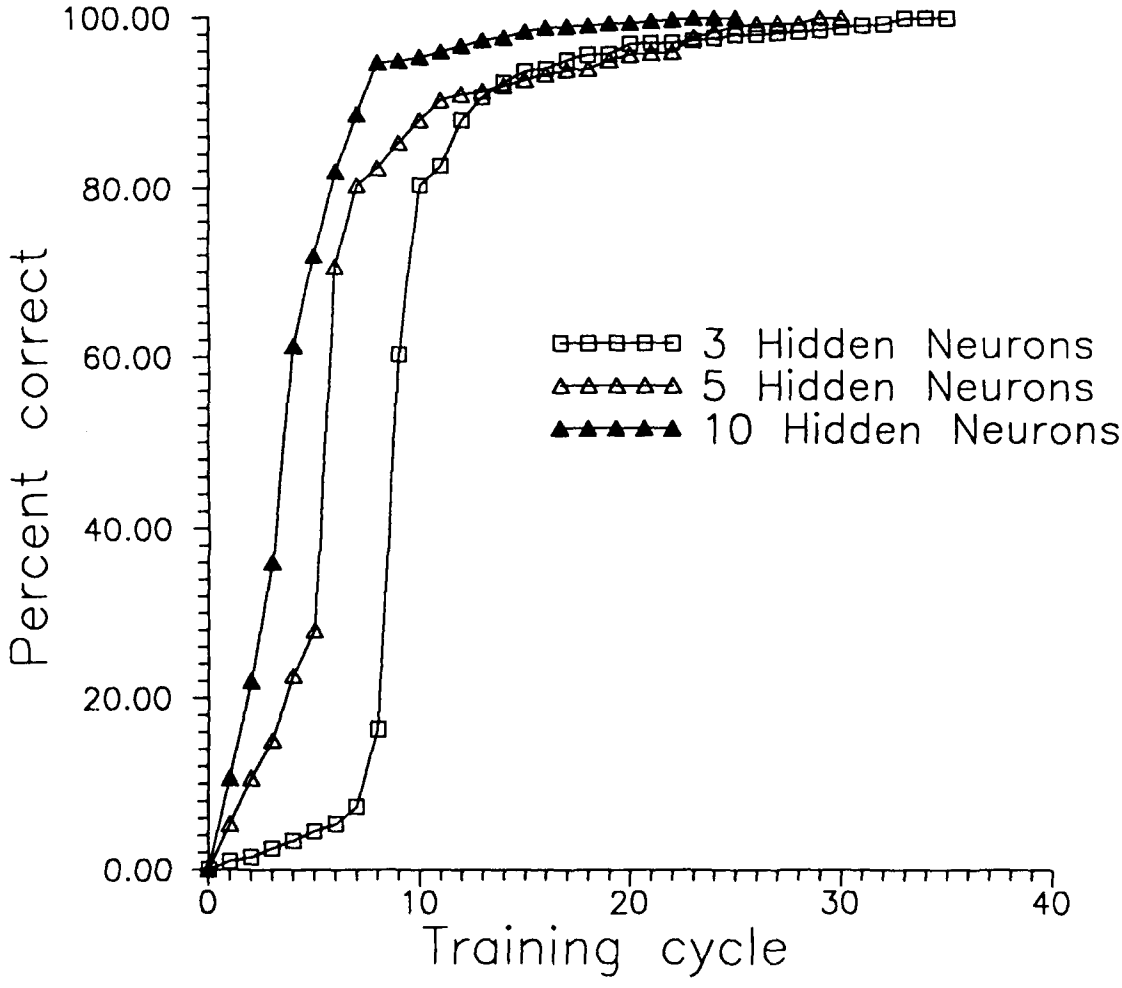


FIGURE 5

Ensemble average learning curves constructed from 20 individual realizations of learning curves for networks with 3, 5, and 10 hidden neurons.

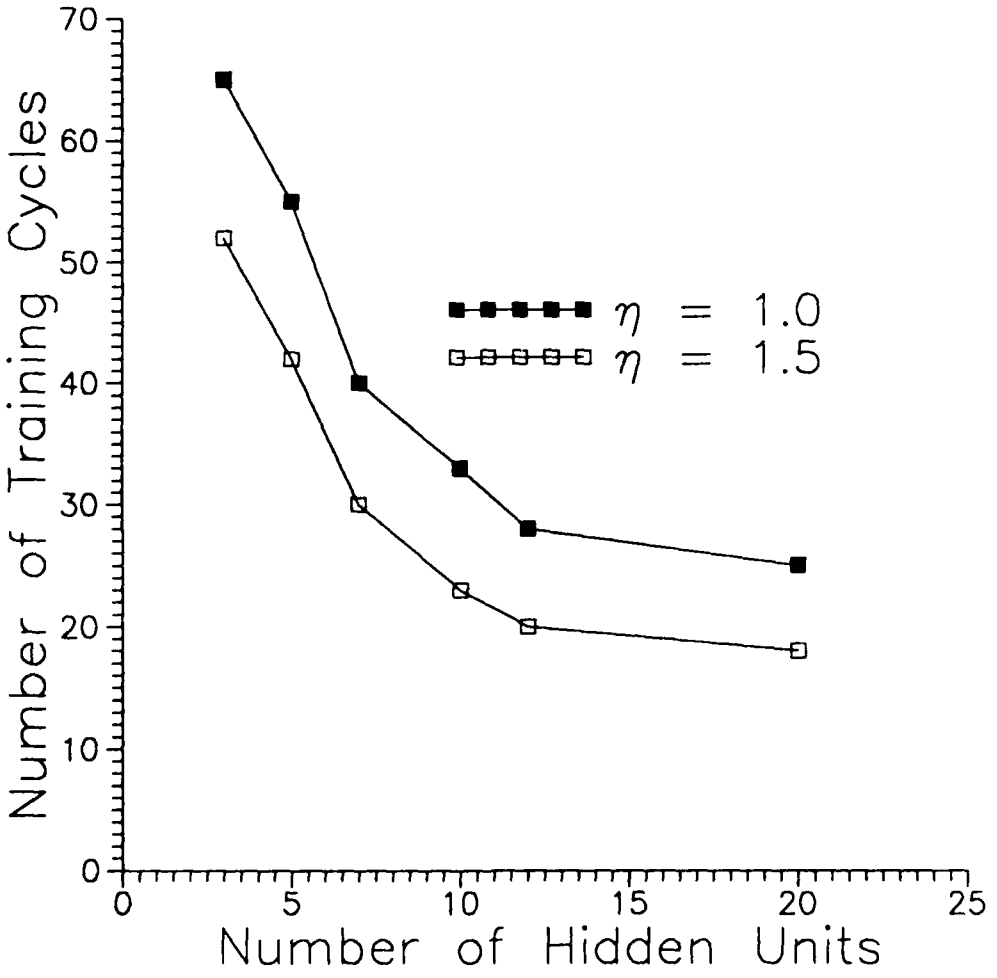


FIGURE 6

Number of training cycles versus the number of hidden units for two values of the learning rate parameter η.

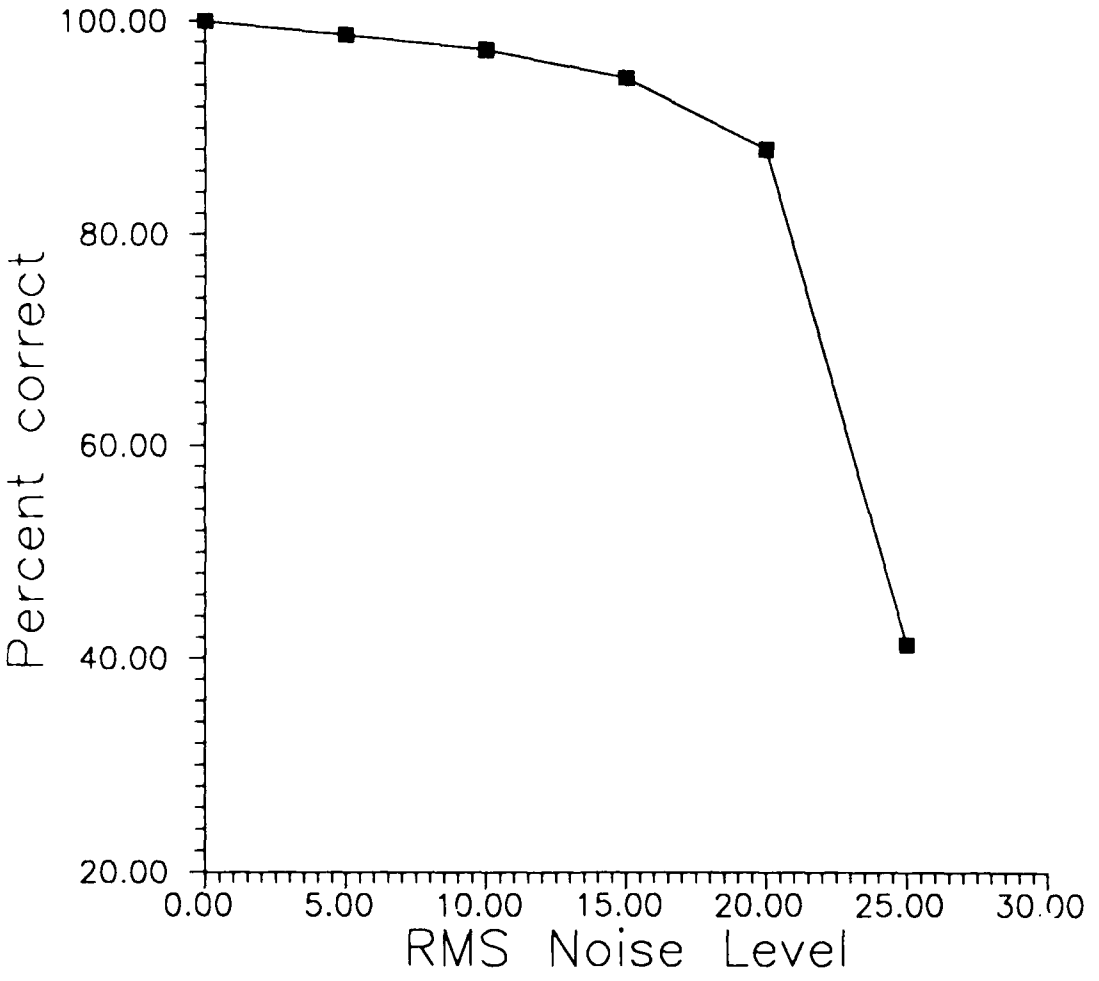


FIGURE 7

Recognition accuracy on the test set versus the RMS level of noise corrupting all weights of a trained network with 20 hidden units.

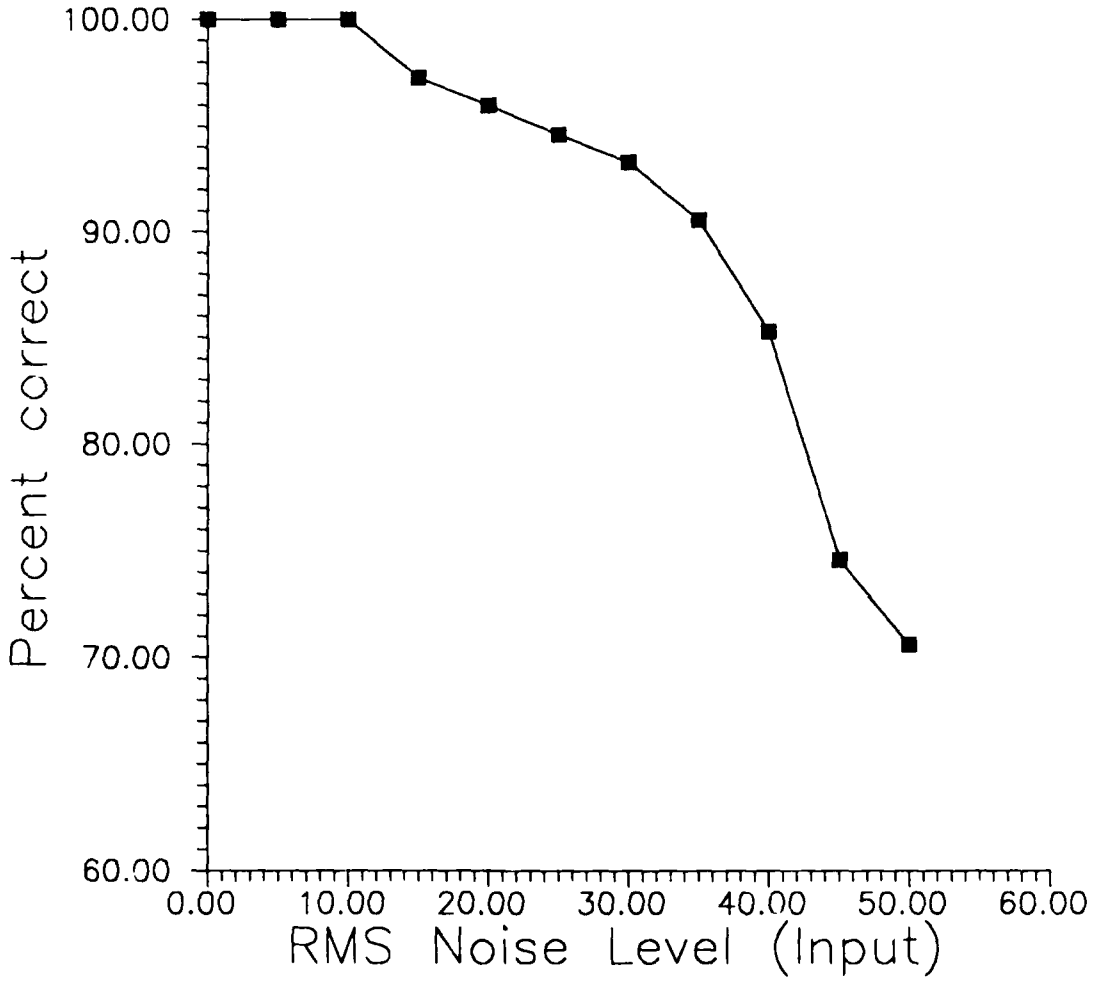


FIGURE 8

Recognition performance of a neural network on the test set. The network was trained on the training set corrupted with a prescribed RMS level of noise.

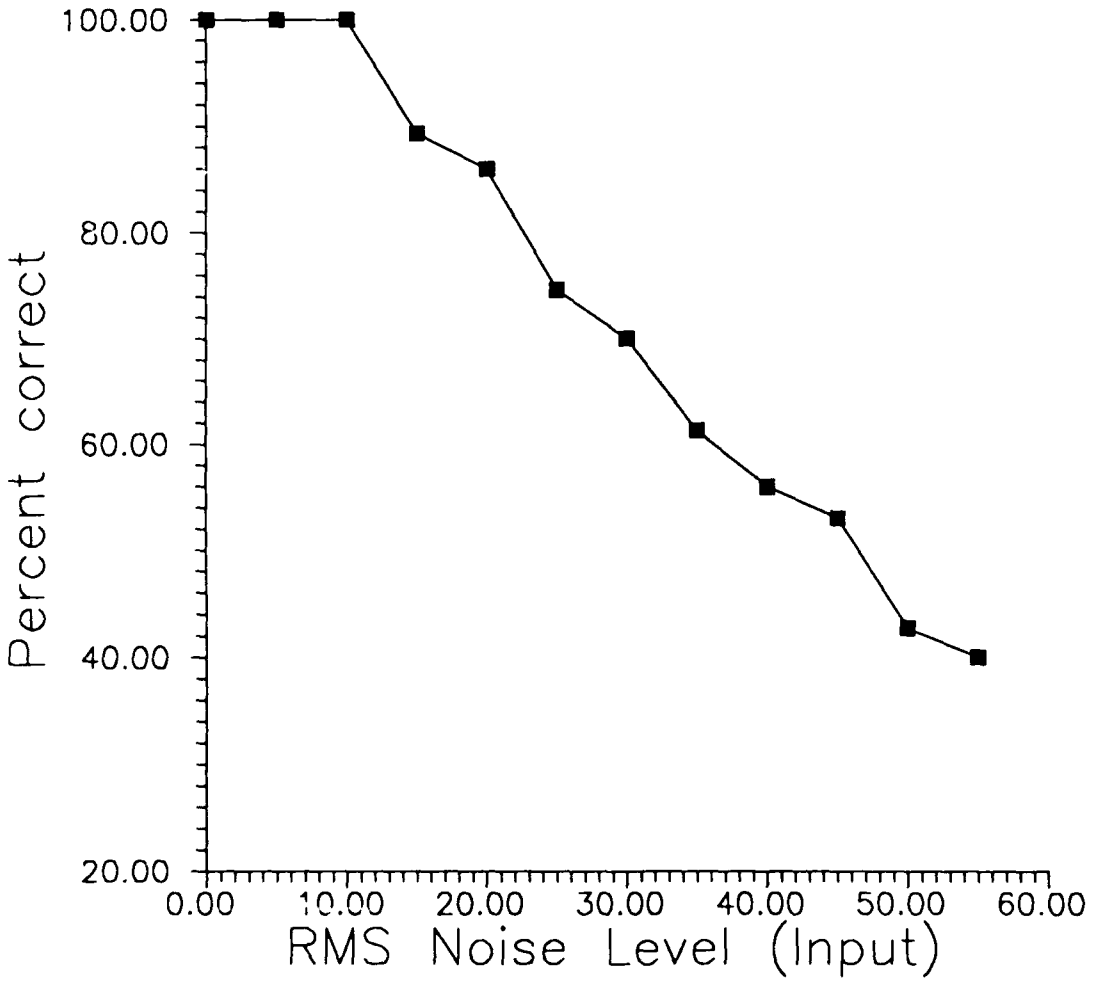


FIGURE 9

Recognition performance of a trained neural network on the test set that has been corrupted with the prescribed RMS level of noise.

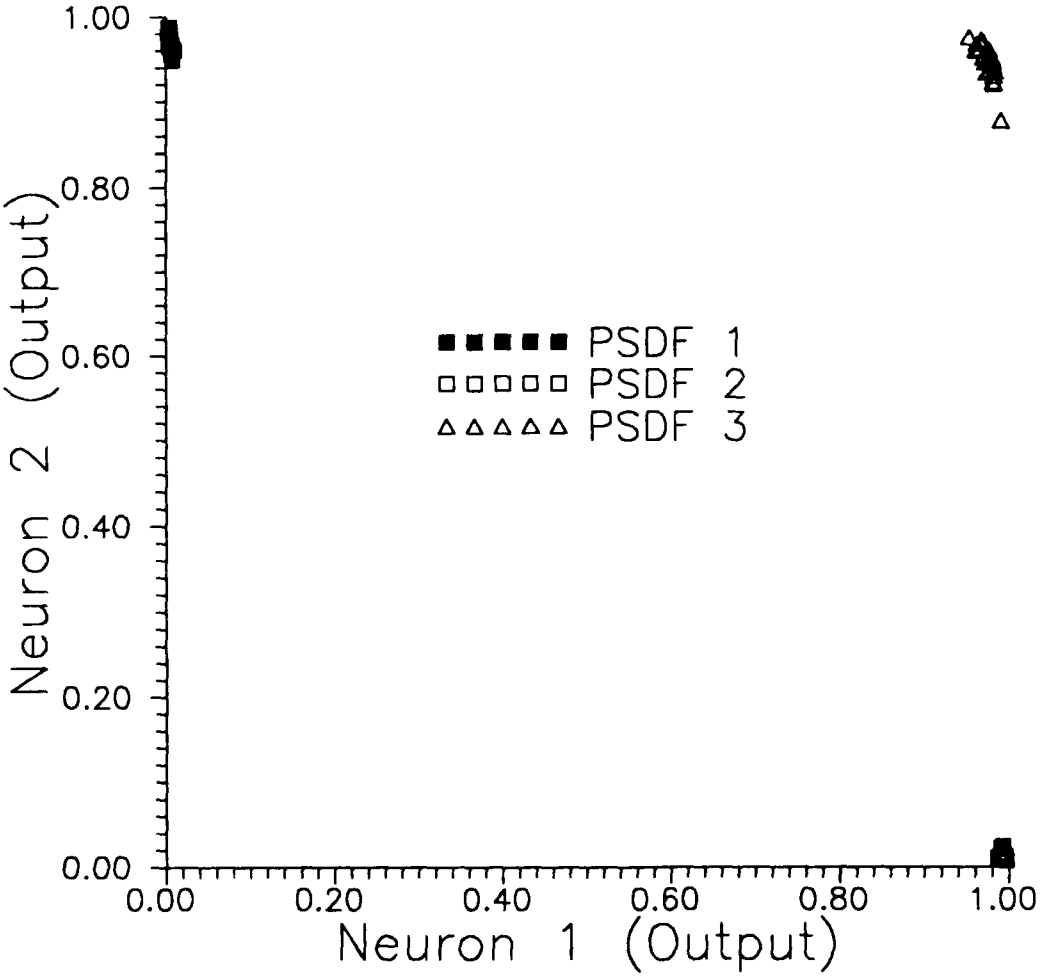


FIGURE 10

Output neural network responses to the test set that has been corrupted at 5 percent RMS noise level.

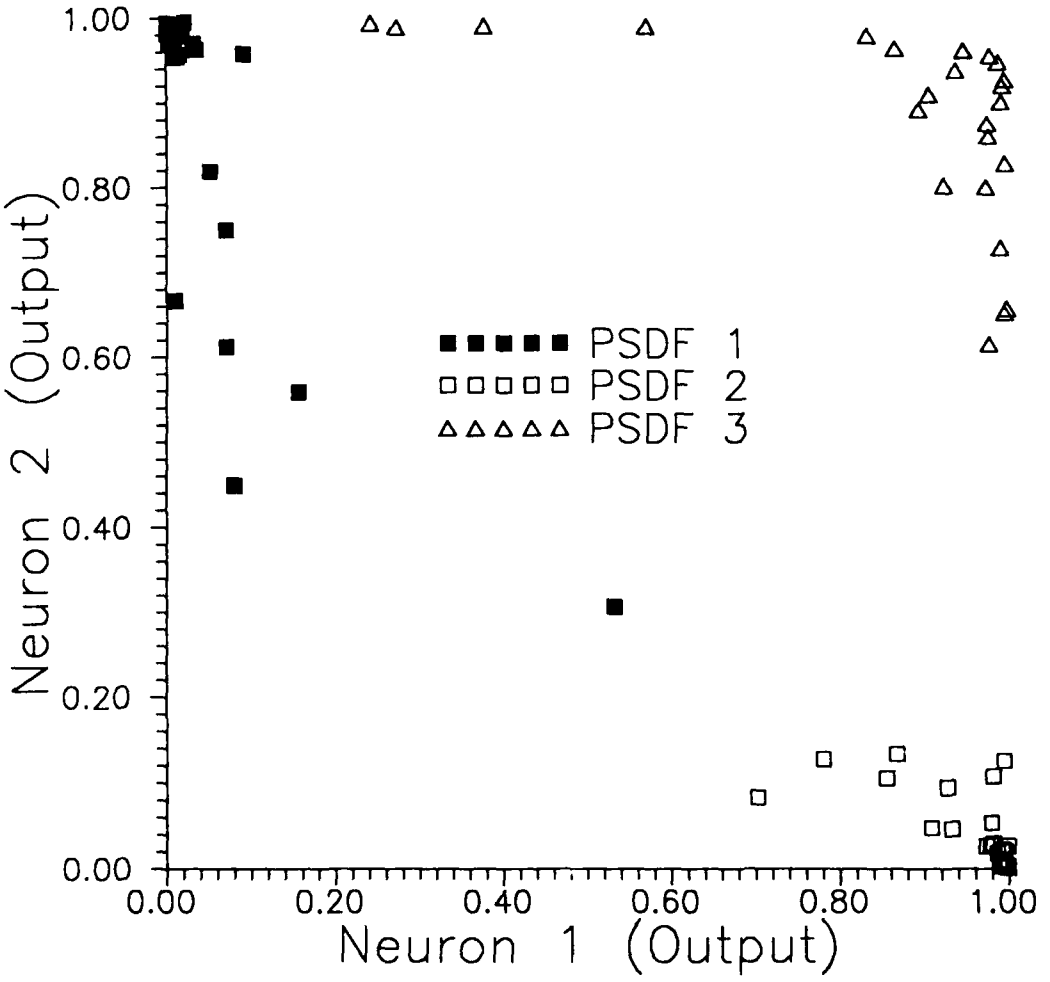


FIGURE 11

Output neural network responses to the test set that has been corrupted at 25 percent RMS noise level.

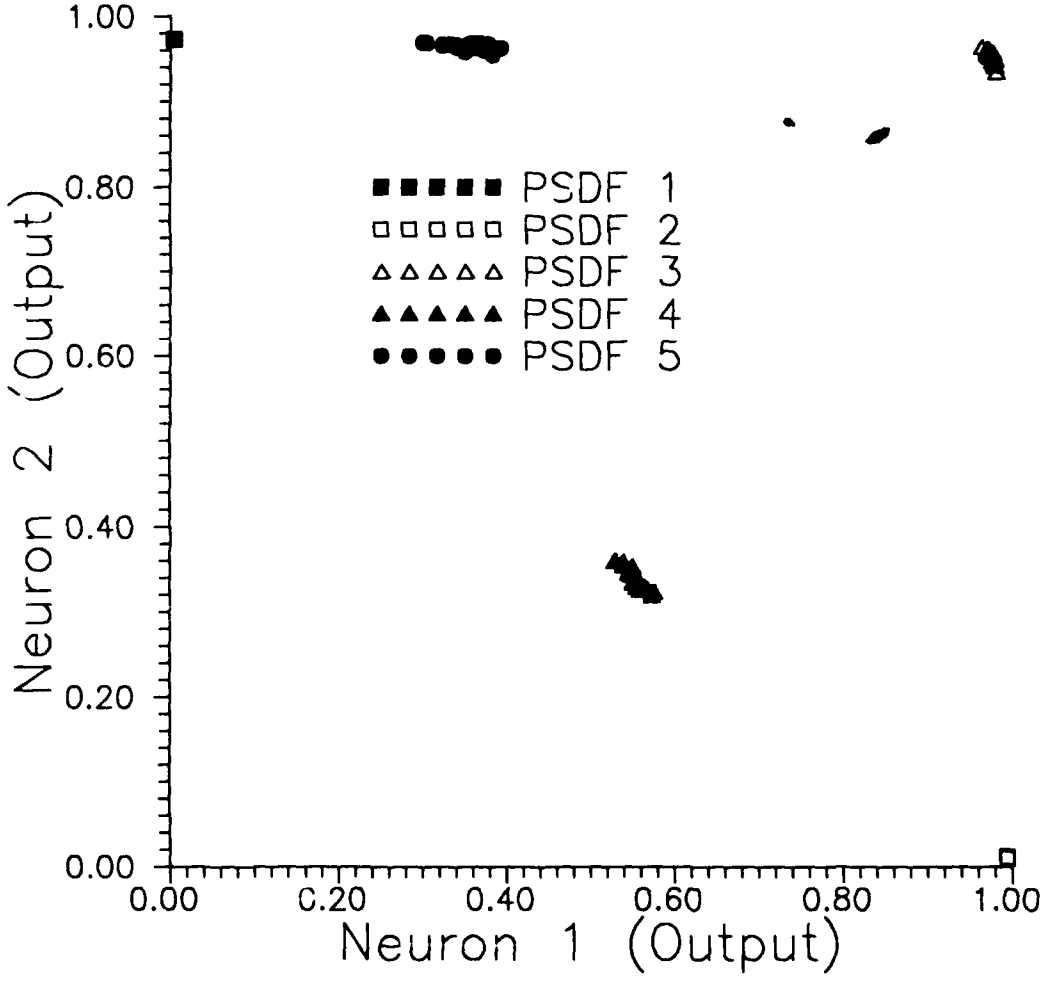


FIGURE 12

Output neural network responses to a data set consisting of a mixture of 25 PSDFs from each of categories 1 to 6. This figure only displays the responses of PSDFs in categories 1 to 5.

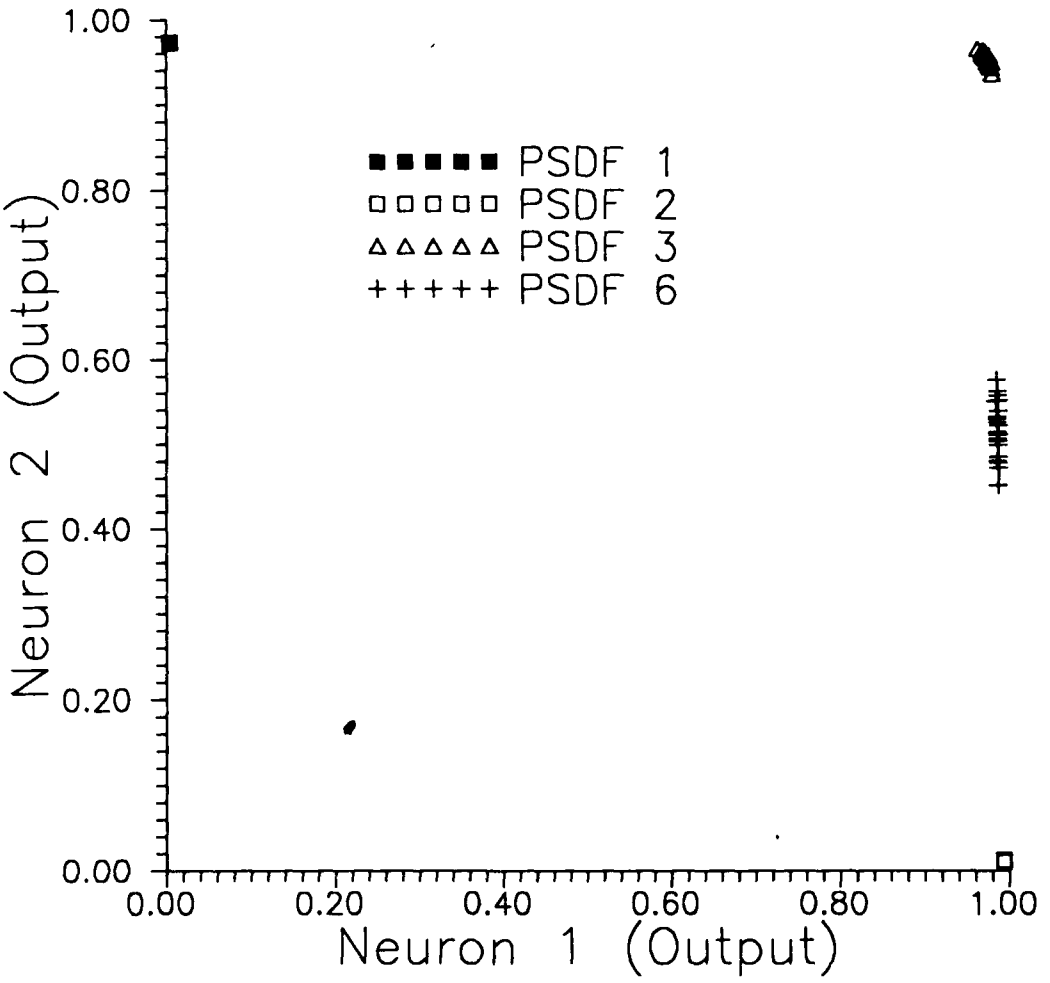


FIGURE 13

Output neural network responses to a data set consisting of a mixture of 25 PSDFs from each of categories 1 to 6. This figure only displays the responses of PSDFs in categories 1, 2, 3, and 6.

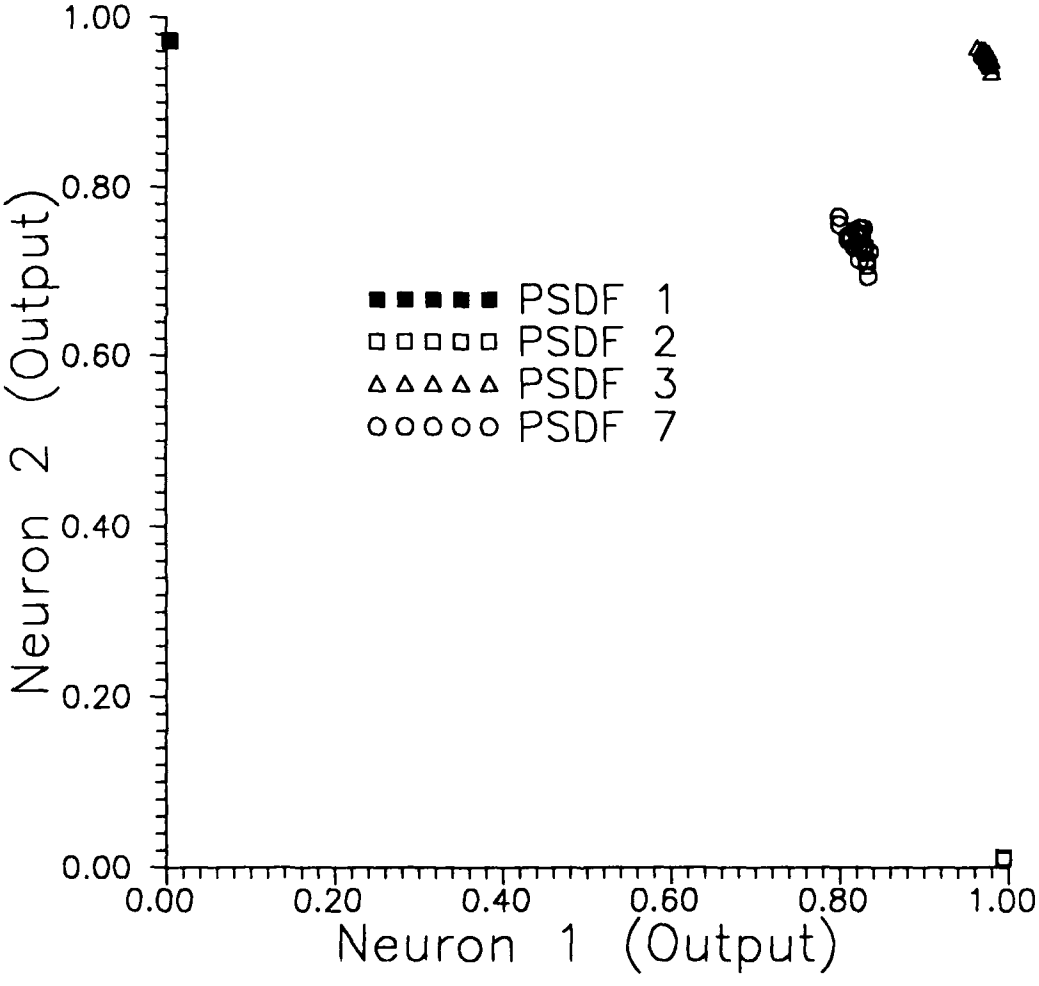


FIGURE 14

Output neural network responses to a data set consisting of a mixture of 25 PSDFs from each of categories 1, 2, 3, and 7.

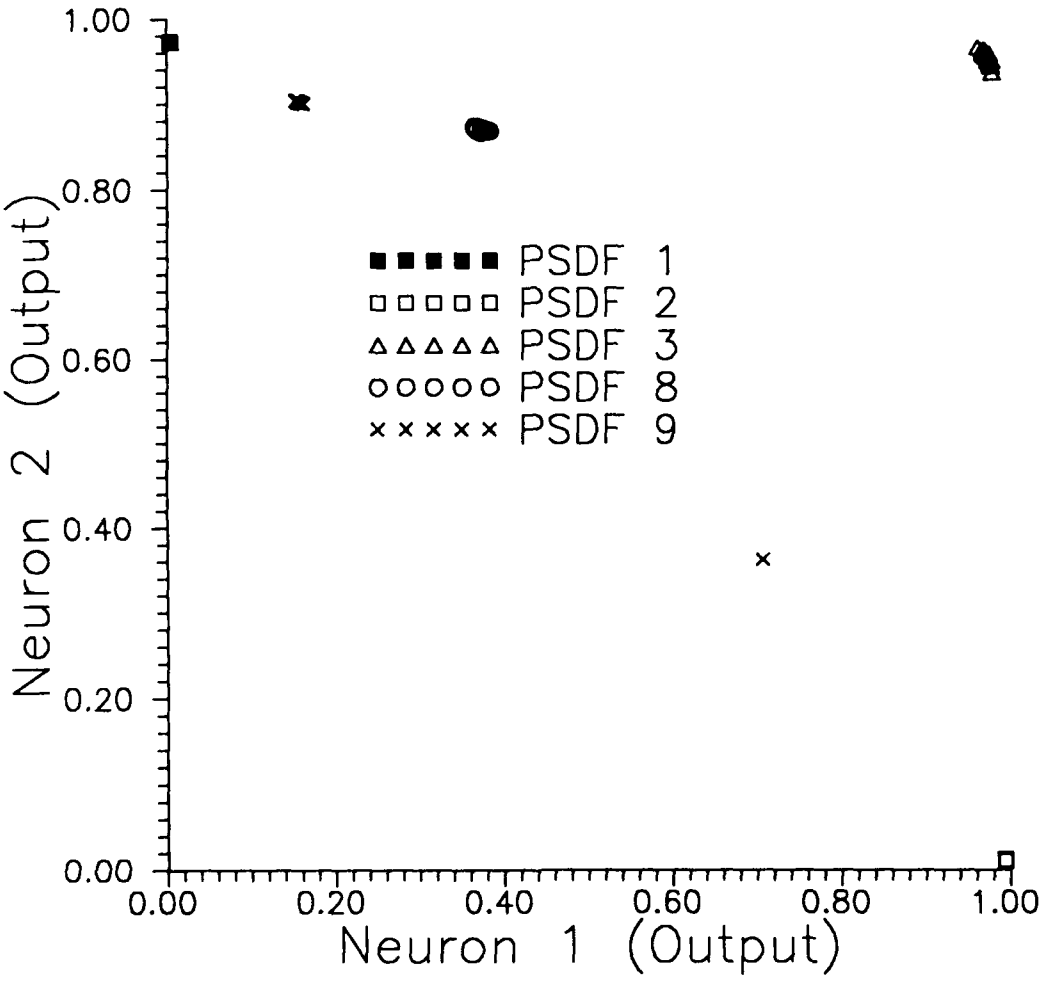


FIGURE 15

Output neural network responses to a data set consisting of a mixture of 25 PSDFs from each of categories 1, 2, 3, 8, and 9.

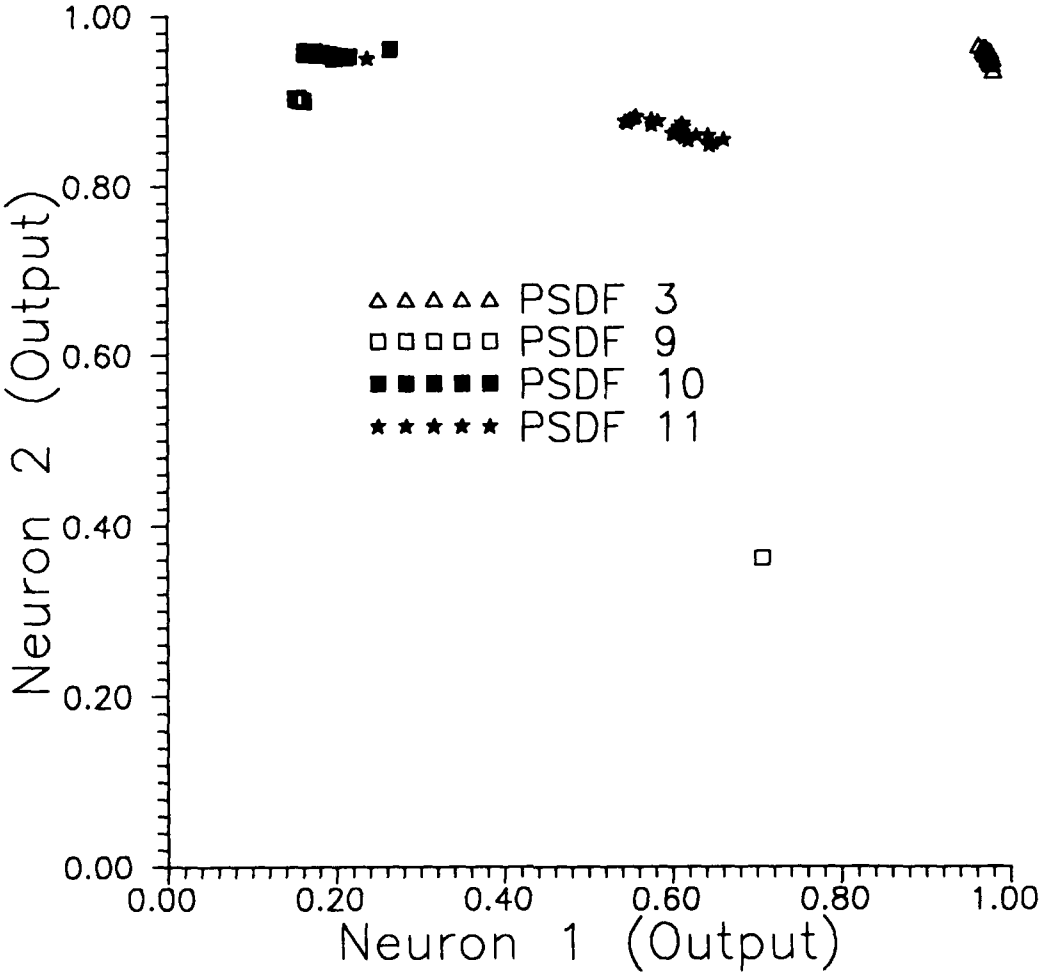


FIGURE 16

Output neural network responses to a data set consisting of a mixture of 25 PSDFs from each of categories 3, 9, 10 and 11.

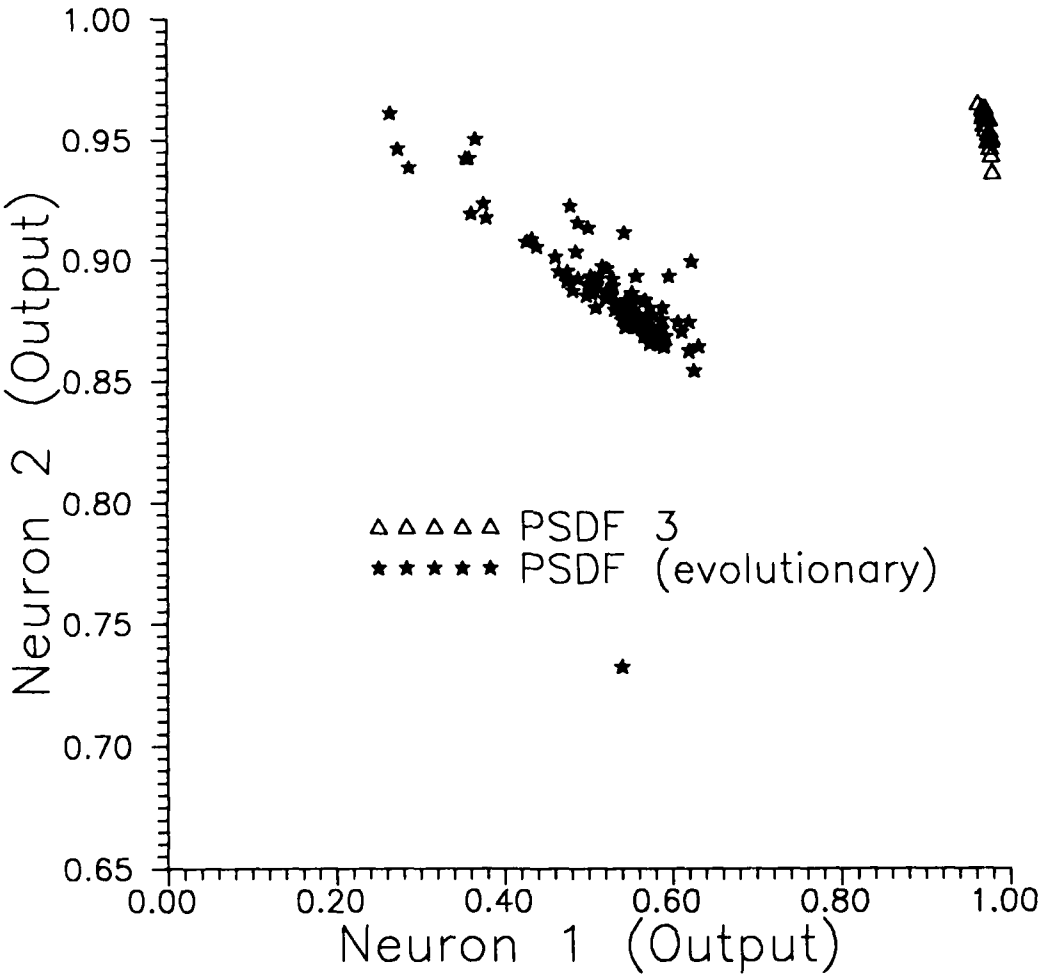


FIGURE 17

Output neural network responses to a data set consisting of a suite of 100 evolutionary PSDFs corresponding to the "spectrogram" from the aerosolization of *Bacillus subtilus var. globigii* in an aerosol chamber and of 25 PSDFs from category 3.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Defence Research Establishment Suffield
Box 4000
Medicine Hat, Alberta, T1A 8K6

2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable)

UNCLASSIFIED

3 TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title.)

NEURAL NETWORK RECOGNITION AND CLASSIFICATION OF AEROSOL PARTICLE SIZE DISTRIBUTIONS (U)

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

YEE, Eugene and HO, Jim

5. DATE OF PUBLICATION (month and year of publication of document)	6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)	6b. NO. OF REFS (total cited in document)
January 1990	34	10

6. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

SR 531 (FINAL) JUNE 1989, OCTOBER 1989

8 SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)
PCN. No. 051SP	

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)
SR 531	

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

- (X) Unlimited distribution
- ( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
- ( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
- ( ) Distribution limited to government departments and agencies; further distribution only as approved
- ( ) Distribution limited to defence departments; further distribution only as approved
- ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

This paper describes the application of a neural computational network model to the pattern recognition and classification of aerodynamic particle size distributions associated with a number of environmental, bacterial, and artificial aerosols. The aerodynamic particle size distributions are measured in real-time with high resolution using a two-spot He-Ne laser velocimeter. The technique employed here for the recognition and classification of aerosols of unknown origin is based on a three-layered neural network that has been trained on a training set consisting of 75 particle size distributions obtained from three distinct types of aerosols. The training of the neural network was accomplished with the back-propagation learning algorithm. The effects of the number of processing units in the hidden layer and the level of noise corrupting the training set, the test set, and the connection weights on the learning rate and classification efficiency of the neural network are studied. The ability of the trained network to generalize from the finite number of size distributions in the training set to unknown size distributions obtained from uncertain and unfamiliar environments is investigated. The approach offers the opportunity of recognizing, classifying, and characterizing aerosol particles in real-time according to their aerodynamic particle size spectrum and its high recognition accuracy shows considerable promise for applications to rapid real-time air monitoring in the areas of occupational health and air pollution standards.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

NEURAL NETWORK
PATTERN RECOGNITION (ADAPTIVE)
AERODYNAMIC PARTICLE SIZE DISTRIBUTION
BACTERIAL AEROSOLS
OPTICAL PARTICLE SIZING
BACK-PROPAGATION LEARNING ALGORITHM
COMMECTIONIST ARCHITECTURES
PARALLEL DISTRIBUTED PROCESSING